ACTI @ EVALITA 2023: Automatic Conspiracy Detection

Giuseppe Russo
ETH Zurich
russog@ethz.ch

Niklas Stoehr
ETH Zurich
nstoehr@ethz.ch

Manoel Horta Ribeiro
EPFL
m.hortaribeiro@epfl.ch

1 Task Description

The ACTI shared task proposes the automatic identification of conspiracy content in Italian language in Telegram. More specifically, it is organized according to two main subtasks:

- **Subtask A: Conspiratorial Content Classification** : a system must recognize if a telegram post is conspiratorial or not.
  - **Conspiratorial**: a text that either (i) expresses the belief that major events(e.g., covid) are manipulation created by powerful people to protect their interests or (ii) interpretation of events meant to contribute to strengthen the underlying narrative of the conspiracy theory.
  - **Not Conspiratorial**: a text that does not diffuse any kind of beliefs linked to conspiracy theory In this task the definition of conspiratorial remain quite broad. Indeed, a sentence is defined conspiratorial even if it shares a claim intended to undermine commonly accepted views on societal issues. For example, the sentence "il cancro femminista sta prendendo piene" is a sentence that should be classified as conspiratorial. As it is subtly supporting a broader theory that claims that women rights are destroying the stability of western societies.
- **Subtask B: Conspiracy Category Classification** : a system must discriminate to which conspiracy theory a post belongs to. In particular, we consider four possible conspiracy theories:
  - **Covid-Conspiracy**: It contains posts concerning vaccine production, 5G , and restrictions as a tool of control over people, and any idea intended to weaken the argument that the pandemic was a real event and actions of people and governments were justified by the seriousness of the issue.
  - **Qanon-Conspiracy**: It contains posts regarding the Qanon-theory according to which a group of Satanic cannibalist sex abusers conspired against former U.S. President Donald Trump during his term in office. The members of this conspiracy has been directly linked to the assault of the Capitol Hill in Washington on January Six. Qanon is a worldwide movement very diffused in Europe (Germany, Italy and Spain mostly). This theory extended far over the original scope embodying other beliefs that support (among the others) the idea that women are enemies (hate against women) and the idea that a powerful elite (led by public figures like Pope Francis, Queen Elizabeth, and Hillary Clinton) is trying to organize a New World Order.
  - **Flat Earth-Conspiracy**: It is a theory claiming that the earth is flat, and there is a great conspiracy supporting the theory that the earth is flat. Usually, the flat-earth conspiracy theory is supported by "scientific evidences".
  - **Pro-Russia Conspiracy**: It is a theory portraying the Russian President Vladimir Putin and Russia as victims of Ukraine and NATO. Such theories is usually supported by claims about the fact that nazists are in charge of Ukraine governments and army.

2 Dataset Description

Our released dataset for **SUBTASK A** training is a csv file containing:

- **id**: It denotes a unique identifier of the post.
- **comment_text**: It represents the text written in the post.
- **conspiratorial**: It is a label that is 1 if the text is conspiratorial and 0 otherwise.

The dataset for the second subtask is a csv file containing:

- **id**: It denotes a unique identifier of the post.
- **comment_text**: It represents the text of a conspiratorial post.
- **conspiracy**: It is a label representing one of the four conspiracy theories indicated in the task description. In particular
  1. **Covid-Conspiracy** has label **0**
  2. **Qanon-Conspiracy** has label **1**
  3. **Flat Earth-Conspiracy** has label **2**
  4. **Pro-Russia-Conspiracy** has label **3**

3 Submission Format

Results for both tasks should be submitted as **csv files**. Submitted runs must contain one result per line including the corresponding **id** field provided in the test sets. In particular for

- **Subtask A: Conspiratorial Content Classification**: The participants that should upload a csv file containing the **id** of the test set samples and the respective predicted label (1 for conspiratorial and 0 for not conspiratorial).
- **Subtask B: Conspiracy Category Classification**: The partocipants that should upload a csv file containing the **id** of the test set samples and the respective predicted label associated with the four conspiracy theories as discussed in the dataset description section.

Additionally, a sample submission for both subtasks will be provided. Only **constrained** runs are accepted.

- **Constrained run**: teams must use only the provided training data from the task organizers

IMPORTANT: Each team can submit up to **FIVE RUNS** for each substask. Further submission will not be considered valid. This constrained might be lifted in the near future and allow for a high number of possible submission.

4 Evaluation

Systems will be evaluated using F1-score (macro) for both subtasks.

We will release the training and the test at the beginning of the competition. To avoid the risk of overfitting on the test set, the rows in the solution file are sampled into Public and Private rows. Hovewer, the partecipants will not know which lines are public or private. As a result, the leaderboard showing the current ranking (available during the entire competition) will use the F1-score on the public rows. We call this **Public Leaderboard**. While the scores obtained on the private rows will be used to the determine the final ranking of the competition. We call this **Private Leaderboard** and the results will only made available at the very end of the competion.

Public rows will consist of roughly 30% of the test data while private rows constitutes the 70% of test data. Therefore, public and private scores may **drastically** change (for unstable submitted systems).

IMPORTANT: THE FINAL SCORE FOR THE COMPETION WILL BE COMPUTED GIVING A WEIGHT OF 60% TO SUBTASK A and of 40% TO SUBTASK B

5 How to Participate and Submit

Partecipate to a Kaggle Competition. ACTI will be hosted on the website Kaggle. Therefore teams that desire to join ACTI must register on Kaggle at this https://www.kaggle.com/account/login?phase=startRegisterTab&returnUrl=%2Fcompetitions. After completing the registration, partecipants should add their email in the Google Form (link on the webpage) to receive the invitation to join the competition.

Submission to ACTI. Particapants should simply click on the invitation link provided. This link will redirect them to a the kaggle webpage where participants can navigate to the tab **Data** to download the data for the competition. **All necessary information are also on the competion links on Kaggle**